AD_____
(Leave blank)

Award Number:
(Enter Army Award number assigned to research, i.e., DAMD17-00-1-0296)
W81XWH-10-1-0499

TITLE:
(Enter title of award)
A Genome-Wide Investigation of Autozygosity and Breast Cancer Risk

PRINCIPAL INVESTIGATOR:
(Enter the name and degree of Principal Investigator and any Associates)

Brandon L. Pierce, Ph.D.

CONTRACTING ORGANIZATION:
(Enter the Name, City, State and Zip Code of the Contracting Organization)

The University of Chicago

Chicago, IL 60637-5418

REPORT DATE:
(Enter month and year, i.e., January 2001)
July 2011

TYPE OF REPORT:
(Enter type of report, i.e., annual, midterm, annual summary, final)

Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: (Check one)

Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
1 July 2011	Annual Summary	7/1/10-6/30/11

4. TITLE AND SUBTITLE

A Genome-Wide Investigation of Autozygosity and Breast Cancer Risk

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-10-1-0499

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Brandon L. Pierce, Ph.D .

bpierce@health.bsd.uchicago.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Chicago,

Chicago, IL 60637-5418

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and
Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Longs segments (> 1 megabase) of homozygous DNA are common in the genomes of outbred human populations. Several lines of research indicate that elevated genome-wide homozygosity and "runs of homozygosity" (RoH) at specific loci may increase breast cancer risk. In this project, we determine if genome-wide RoH content and individual RoHs are more common in breast cancer cases than in controls, using logistic regression methods. **U**sing genome-wide SNP data (525,000 SNPs) on 1,647 non-Hispanic white, early-onset premenopausal breast cancer cases and 1,556 matched controls we identified over 65,000 individual RoHs and 423 genomic regions harbor RoHs for at least 10 individuals in our dataset (i.e., "RoH regions"). Overall RoH content was not associated with breast cancer status or with subtypes of breast cancer as defined by estrogen receptor status. Furthermore, analyses of each of the 423 RoH regions did not reveal any region in which RoH status was significantly associated with breast cancer risk or risk for a breast cancer subtype (after correction for multiple testing). Finally, comparing the RoH regions showing the strongest associations in our study to the regions with the strongest association in a prior study of RoHs in breast cancer did not reveal any common findings across studies. In this association study of RoHs and early-onset breast cancer risk, we have not implicated overall RoH content or any individual RoH in breast cancer risk.

15. SUBJECT TERMS
Runs of Homozygosity, Autozygosity, early-onset breast cancer, genome-wide association, single nucludotide polymorphism

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU	39	USAMRMC

Table of Contents

Page

INTRODUCTION

Longs segments (> 1 megabase) of homozygous DNA are common in the genomes of outbred human populations (1-5). Several lines of research indicate that elevated genome-wide homozygosity (i.e. autozygosity) may increase breast cancer risk (6-9). For several cancer types, small studies have found increased germline homozygosity at specific genomic locations, suggesting these regions harbor important cancer genes (10, 11). Homozygosity mapping is a natural extension of large genome-wide association studies and has the potential to identify novel breast cancer genes and provide biological insights. Based on this evidence, we hypothesize that germline autozygosity is more common in breast cancer cases than in controls. More specifically, we hypothesize that there are specific regions of the genome in which homozygosity (i.e. "runs of homozygosity" (RoHs)) are more common in breast cancer cases than in controls and that these regions contain breast cancer-related genes.

BODY

A recent presentation prepared to the Era of Hope meeting in August 2011 (Orlando, FL) provides an overview of the progress this project to date. The presentation file is attached. Slides of this presentation are referenced throughout this report.

Description of progress towards accomplishing tasks described in scope of work document:

Task 1 data acquisition and preparation (Slide 7): We have obtained all the genotype data needed for this project and performed all standard quality control procedures (removal of population outliers and samples of poor quality; SNP filtering based on call rates, Hardy-Weinberg Equilibrium, allele frequency; assessment of population structure).

Task 2: genome-wide autozygosity analysis (Slides 8 -18): We have identified runs of homozygosity (RoH) in our data using two different methods, as implemented in the Golden Helix and PLINK programs, respectively. We have used these RoHs to derive genome-wide measures of overall homozygosity. We have tested these measures for association with case/control status and also performed sub-group analyses by estrogen receptor status.

Task 3: Conduct autozygosity mapping analysis (Slides 19-22): using the Golden Helix data on genome-wide RoHs, we have identified 423 regions in which homozygosity is somewhat common (>10 occurrences in our dataset).

Task 4: CGEMS analysis: RoH analysis in the publically available CGEMS dataset has recently been published by another group (12), and we have compared our findings to this work (see below) Instead, we will follow-up on our findings using an additional set of xxx early-onset breast cancer cases and xxx controls typed on the Cyto12 300K SNP chip.

Task 5: DNA Sequencing: We are further evaluating our findings in order to determine a promising region (or regions) for DNA sequencing

Connection to Previous findings (Slide 26): Based on our preliminary findings, we find no overlap between our most significant RoH regions and the regions identified in a previous analysis of RoHs in a case-control study of breast cancer. This prior study focused primarily on ER+ breast cancer, we restricted this comparison to our ER+ results.

Future Work: Several additional analyses will allow us to further confirm our results. First, we will conduct RoH mapping in an independent set of early-onset breast cancer cases and controls. This dataset is of comparable size, but is typed on a different Illumina chip (Cyto12: 300,000 SNPs). We will also systematically compare out Golden Helix RoHs with our PLINK

RoHs to determine if the RoHs identified vary substantially by algorithm used. We will also examine probe intensity data for RoH segments of interest to confirm that each RoH is due to homozygosity and not a deletion. Finally, regarding our proposed sequencing of regions identified in RoH analyses, we will perform such sequencing should a promising gene region emerge from these analyses. At this point, we have not identified such a region.

KEY RESEARCH ACCOMPLISHMENTS

- Obtaining and performing quality control procedures on GWAS data
- Estimation and description of RoHs estimated from GWAS data.
- Test of runs of homozygosity (both genome-wide RoH levels and individual loci) for association with early-onset breast cancer

REPORTABLE OUTCOMES

- An abstract describing this work has been accepted for platform and poster presentation to the Era of Hope meeting to be held in August 2011 in Orlando, FL.
- This Award has supported the post-doctoral training that helped the P.I. receive several job offers for tenure-track faculty positions.

CONCLUSION (Slide 27):

In this work, we find no evidence that overall RoH content or specific RoHs contribute to early-onset breast cancer risk. However, our most promising RoHs will be followed-up in an independent sample of non-Hispanic white early-onset breast cancer cases and controls. We found some suggestive association for total Mb of RoH and ER+ breast cancer and this finding will also have to be followed-up in an independent sample.

The utility of RoH analysis for detection of cancer susceptibility loci appears to be somewhat limited, at least in case-control GWA studies using standard SNP panels measured on a few thousand individuals. The technique may be more appropriate in populations with higher RoH content. However, with available of large GWAS datasets the lack of association for RoHs can be easily confirmed for other cancer phenotypes.

REFERENCES

1. Broman KW, Weber JL. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. Am J Hum Genet 1999;65:1493-500.
2. Curtis D, Vine AE, Knight J. Study of regions of extended homozygosity provides a powerful method to explore haplotype structure of human populations. Ann Hum Genet 2008;72:261-78.
3. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. Hum Mol Genet 2006;15:789-95.
4. Li LH, Ho SF, Chen CH, et al. Long contiguous stretches of homozygosity in the human genome. Hum Mutat 2006;27:1115-21.

5. Simon-Sanchez J, Scholz S, Fung HC, et al. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. Hum Mol Genet 2007;16:1-14.
6. Rudan I, Rudan D, Campbell H, et al. Inbreeding and risk of late onset complex disease. J Med Genet 2003;40:925-32.
7. Shami SA, Qaisar R, Bittles AH. Consanguinity and adult morbidity in Pakistan. Lancet 1991;338:954.
8. Gilani GM, Kamal S. Risk factors for breast cancer in Pakistani women aged less than 45 years. Ann Hum Biol 2004;31:398-407.
9. Rudan I. Inbreeding and cancer incidence in human isolates. Hum Biol 1999;71:173-87.
10. Bacolod MD, Schemmann GS, Wang S, et al. The signatures of autozygosity among patients with colorectal cancer. Cancer Res 2008;68:2610-21.
11. Assie G, LaFramboise T, Platzer P, Eng C. Frequency of germline genomic homozygosity associated with cancer cases. JAMA 2008;299:1437-45.
12. Enciso-Mora V, Hosking FJ, Houlston RS. Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. Eur J Hum Genet;18:909-14.

APPENDICES

A copy of a the slides from a recent presentation of this research is attached. The PI has also attached his CV.

SUPPORTING DATA

None

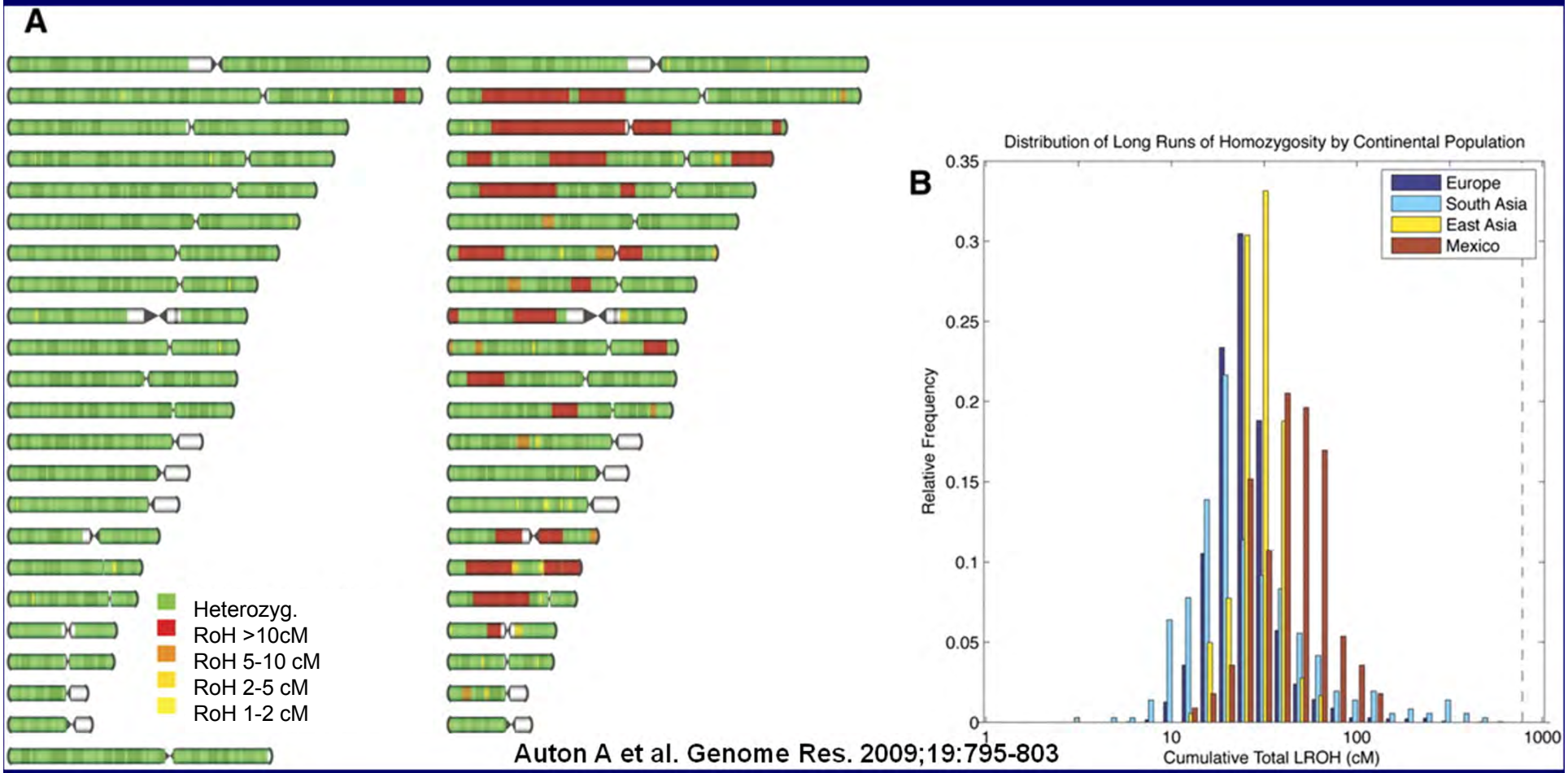
Runs of Homozygosity (RoH) and Early-Onset Breast Cancer Risk

Brandon L. Pierce
University of Chicago
August 4, 2011

Runs of Homozygosity (RoH)

- **RoH**: an long segment of consecutive homozygous genotypes (~1 Mb)
  - Suggesting the chromosome pair share an identical segment
  - Relatively common in human genomes
  - Distribution of RoH frequency and size varies by population
- RoHs could occur due to
  - Related parents (close or distant) (i.e., “autozygosity” or IBD)
  - Natural selection
  - Hemizygosity (e.g. deletion)
  - Uniparental isodisomy
- RoHs have been used to map Mendelian disease genes

Runs of Homozygosity in Humans



Auton A et al. Genome Res. 2009;19:795-803

Rationale for RoH as a breast cancer risk factor

- Some studies show consanguinity as a risk factor (human and mice)
  - Reflecting increased RoH content
- Loss of homozygosity (LOH) is a common event in tumors
  - In a similar fashion could RoH influence tumor formation?
- RoH could harbor recessive susceptibility variants
  - Not easily detectable in GWAS or in linkage studies?
  - GWAS have found very few variants with clear recessive effects
  - RoH mapping has linked recessive loci to schizophrenia risk
- RoHs have been used to map Mendelian disease genes
- Breast cancer susceptibility
  - Common variants (e.g., FGFR2), rare variants of strong effect (e.g., BRCA1/2), and variants of intermediate frequency (e.g. CHEK2)

Studies of RoH and Cancer Risk

- Initial Studies (positive findings)
  - Bacoloc et al (2008): RoH more common in CRC cases (50K SNPs)
  - Assie et al. (2008): RoH at specific loci more common for breast, prostate and head/neck (345 micro-satellite markers)
- Colorectal cancer
  - Spain et al (2009): no replication of Bacoloc et al (550K SNPs)
- Acute lymphoblastic leukemia (Hosking, 2010)
  - Homozygosity unlikely to affect risk (292K SNPs)
- Breast and prostate cancer (Enciso-Mora 2010)
  - No strong evidence that homozygosity increases risk (550K SNPs)
- Early-onset BrCa not yet studied*

Overall Goal

- Determine if RoHs are related to early-onset breast cancer risk
- Hypotheses:
  - Overall gernlime RoH is more common in breast cancer cases than controls
  - Homozygosity at specific genomic regions is more common in cases than controls
- Implications:
  - Such regions are likely to harbor cancer-related genes

Early-onset breast cancer GWAS

- 3,203 non-Hispanic white participants
  - 1,647cases , 1,556 controls
  - From BCFR (California, Ontario), Germany, Seattle, Long Island
  - Known BRCA1 and BRCA2 carriers excluded
- Typed for ~610,000 SNPs using Illumina 610-Quad chip at Univ. of Chicago
- Standard GWAS QC filters based on call rates (>0.97), allele frequency (>0.05), and HWE (p>0.0001)
- Principle components analysis for
  - Exclusion of individual of non-European ancestry
  - Adjustment for variation in European ancestry

Methods for RoH Detection

- Golden Helix RoH module
  - detects RoHs according to:
    - A minimum number of consecutive homozygous SNPs (i.e., 100 SNPs) allowing for error and
    - A minimum physical distance (i.e., 1 Mb)
  - Creates “clusters” of runs for association testing
    - Defined as a continuous set of SNPs were each SNP has at least 10 samples with an RoH within that set
- PLINK ROH command
  - A “sliding window” approach

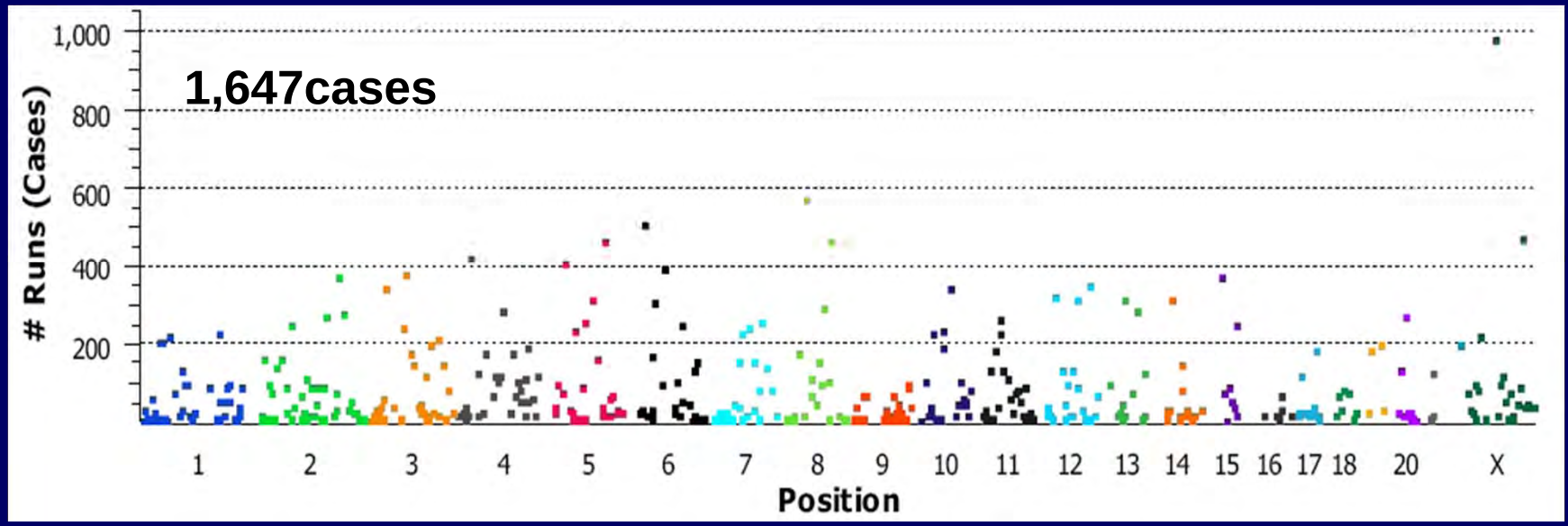
Methods for RoH Detection

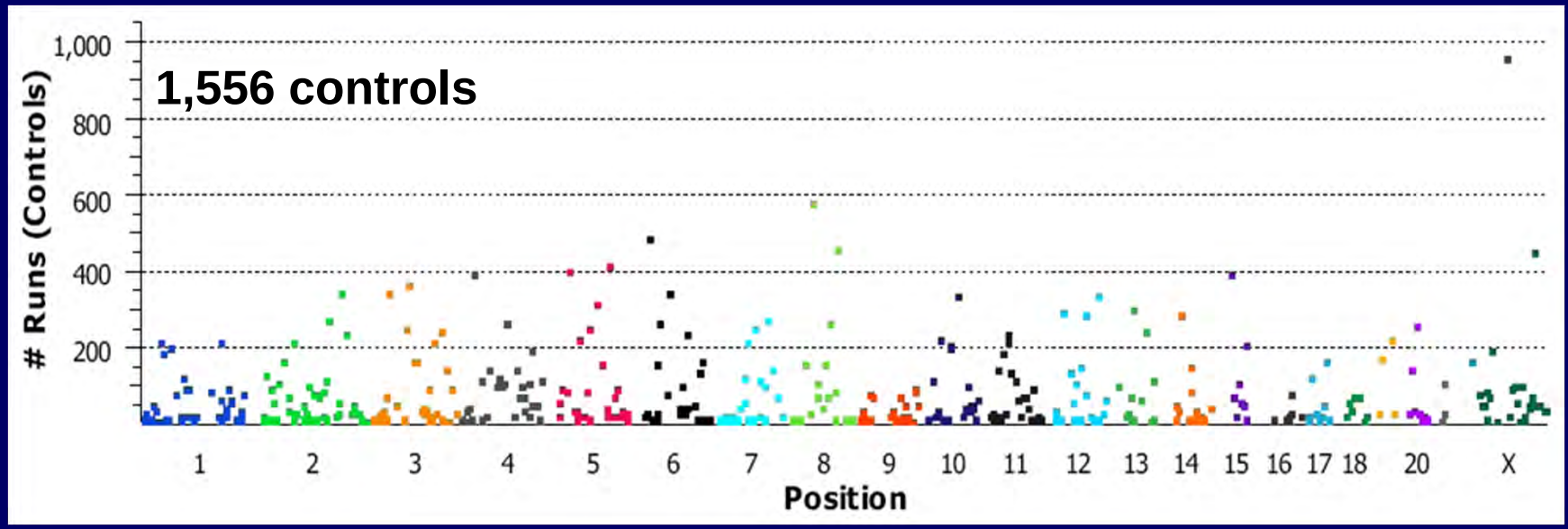
- Full Panel of SNPs (525K post-QC SNPs)
  - Avg. density = 1SNP/5kb
  - 310K tagSNPs or "haplo-groups" (at r^2 of 0.7)
    - Reducing information by ~40%
- We need to detect RoHs with high confidence!
  - If SNPs were independent, a randomly generated RoH would occur with probability:
    - $(1\text{-}0.34)^{(\#SNPs)}$ x 525,000 SNPs x 3,203 participants
    - Lencz, et al 2007
  - For a 0.05 probability, a run of ~60 SNPs is required.
  - Due to 40% reduction in info: ~100 SNPs is required (2% error allowed)
- Minimum length of 1Mb
  - To eliminates shorter RoHs in SNP-dense, high-LD areas

RoH Detection Results

- Total RoH segments detected across all individuals:
  - Golden Helix method: 66,633 detected
  - PLINK method: 68,335 detected
- Mean length of RoH:
  - GH: 1.44 MB and 187 SNPs
  - PLINK : 1.65 Mb and 194 SNPs
- Number of commons RoH clusters or pools:
  - GH: 428
  - PLINK: N/A

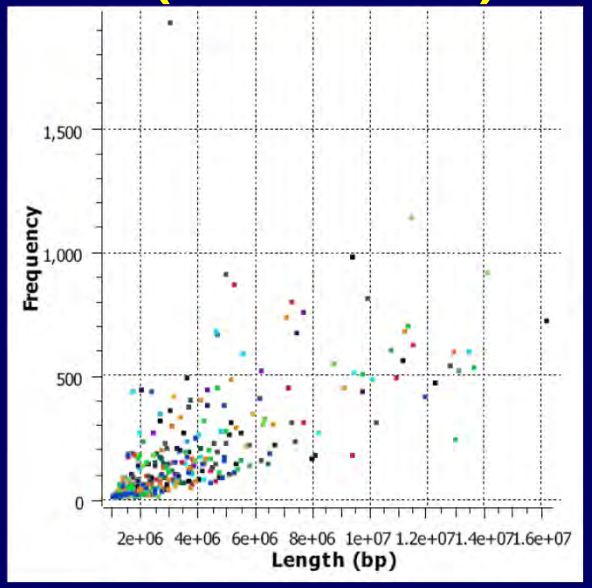
Frequency of Common RoHs (n=423)

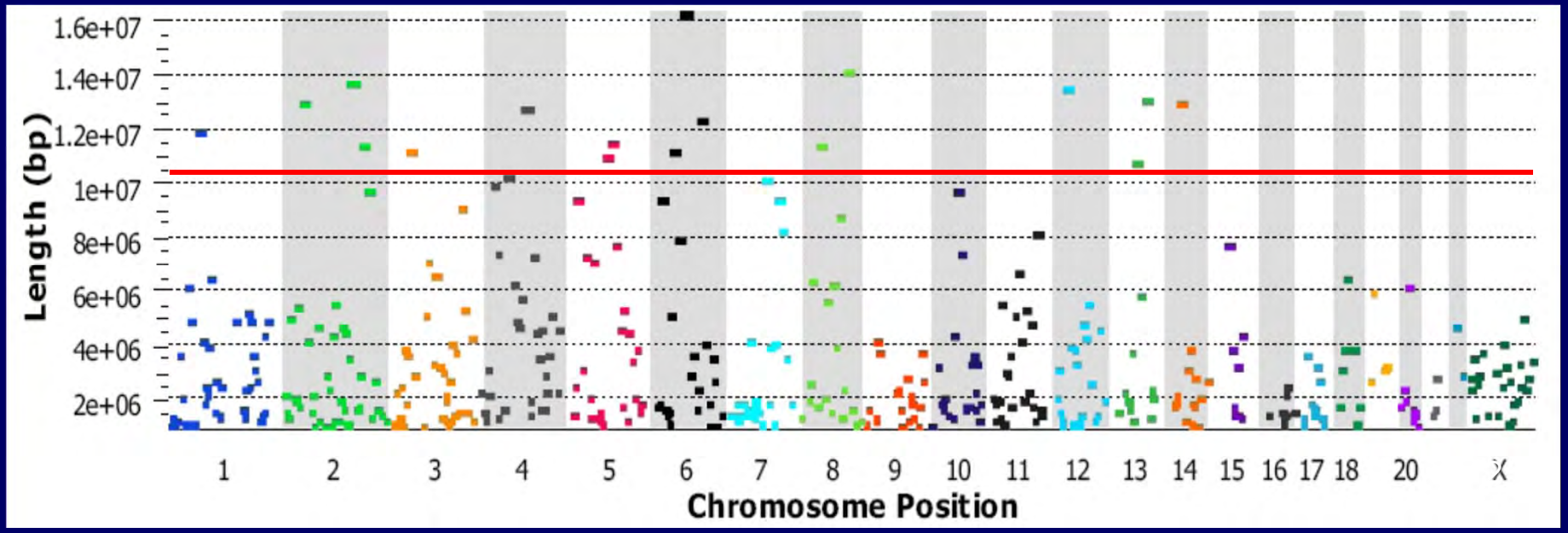




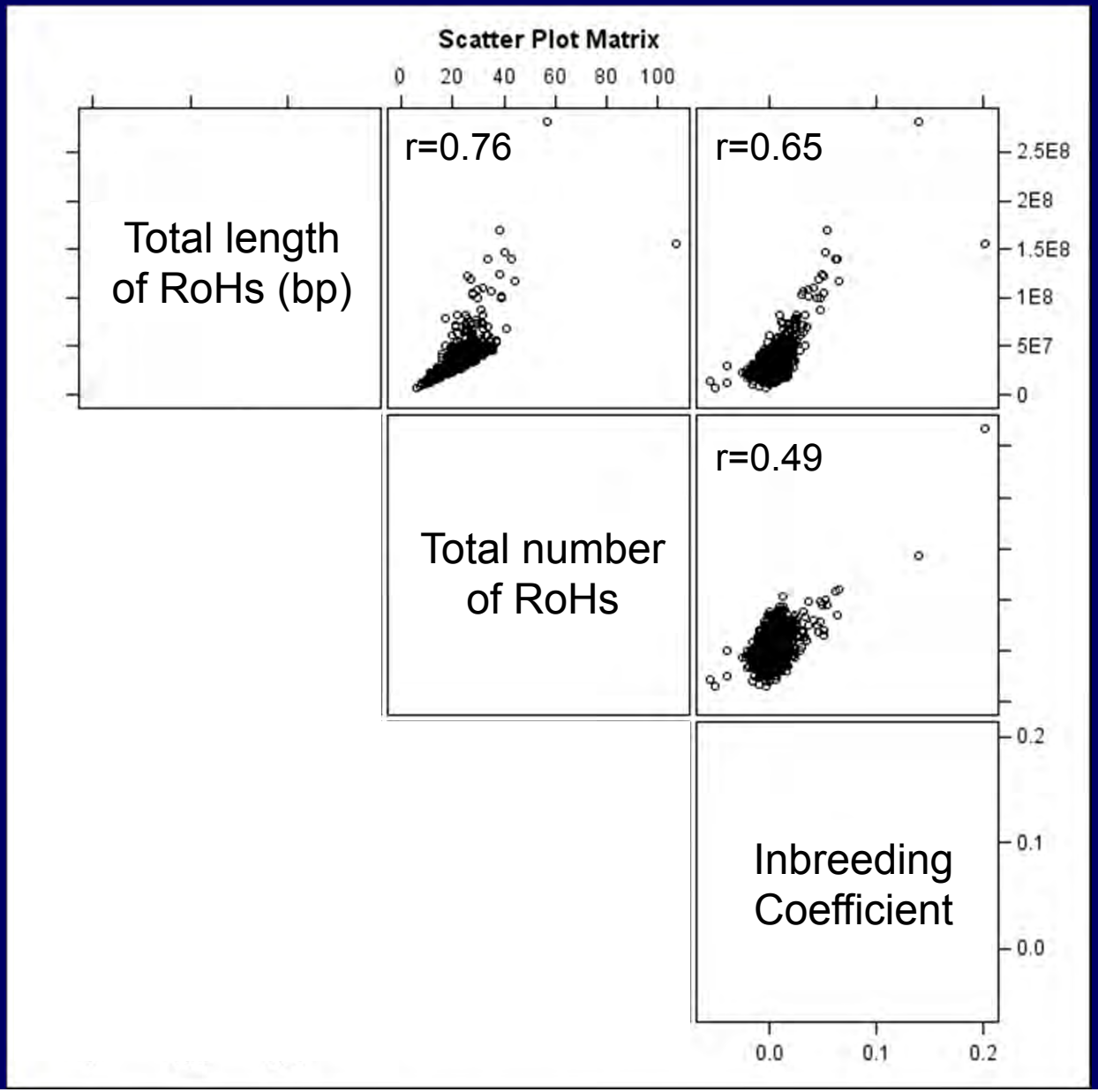
Length of “Common” RoHs (n=423)

- 19 RoH >10Mb, Including:
  - various centromeric regions (e.g. 6, 8, 12)



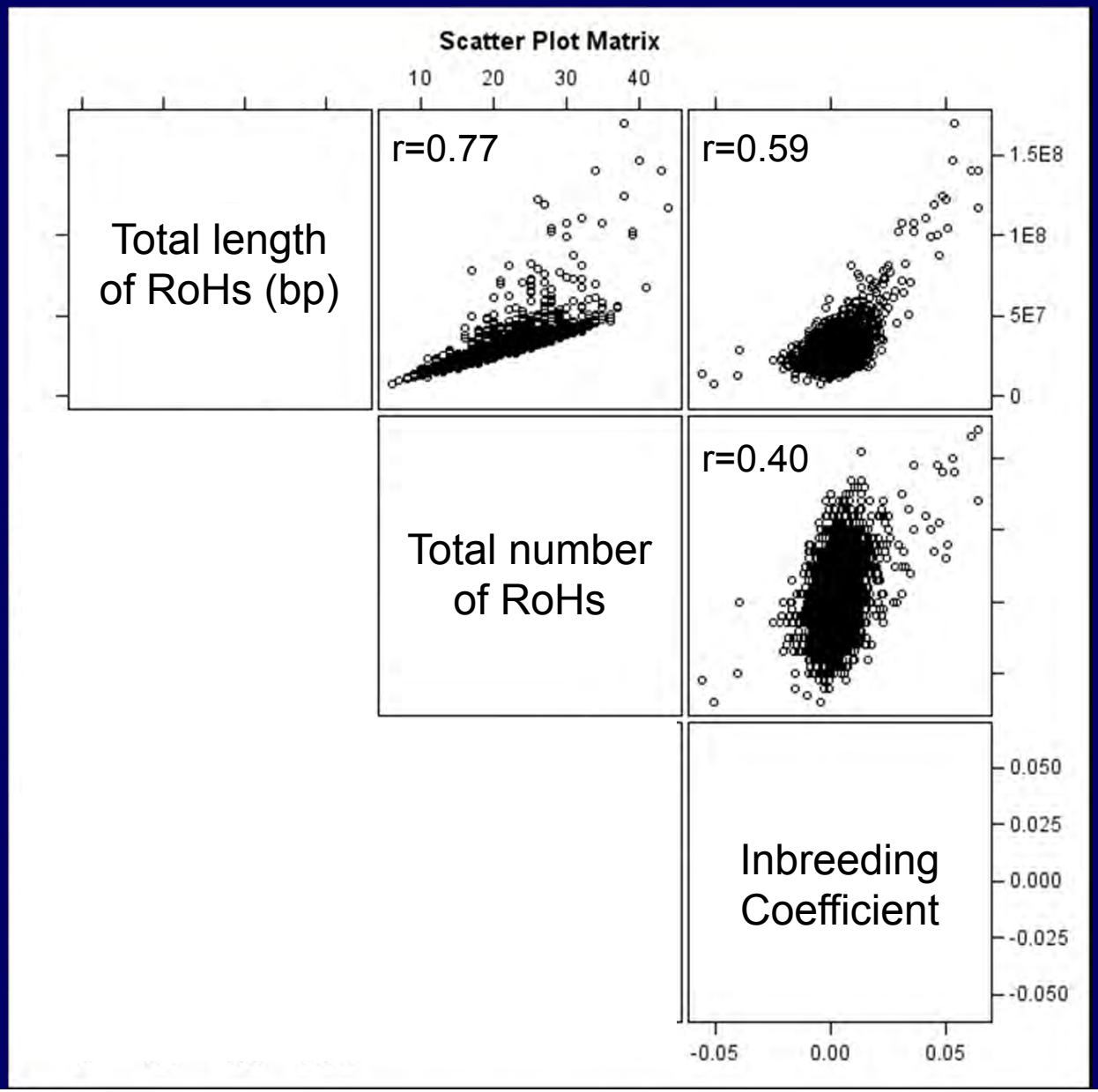


Correlations b/t Consanguinity Measures (blcok out)



Correlations b/t Consanguinity Measures

(excluding outliers)



Association analyses

- Test association between breast cancer status and
  - Total number of RoHs per individual
  - Total length of RoHs per individual
- Using logistic regression adjusted for
  - Age at diagnosis/interview
  - PCA-derived ancestry (5 PCs)
- Examine by ER status

Association between overall RoH and BrCa risk (1641 cases; 1554 controls)

Number of RoHs	OR	95% CI	P
6-18	1.00	Ref	
19-21	0.81	0.68-0.99	0.03
22-24	0.94	0.77-1.14	0.50
>25	0.97	0.79-1.18	0.74
Total Length (Mb)			
7.5-24.1	1.00	Ref	
24.2-28.6	0.82	0.67-1.00	0.05
28.7-33.3	0.80	0.66-0.98	0.03
>33.5	0.96	0.79-1.17	0.69

Association between overall RoH and ER+ risk (972 cases; 1554 controls)

Number of RoHs	OR	95% CI	P
6-18	1.00	Ref	
19-21	0.85	0.98-1.05	0.13
22-24	1.05	0.84-1.31	0.66
>25	0.98	0.77-1.23	0.84
Total Length (Mb)			
7.5-24.1	1.00	Ref	
24.2-28.6	0.91	0.72-1.15	0.42
28.7-33.3	0.87	0.69-1.10	0.25
>33.5	1.00	0.80-1.27	0.94

Association between overall RoH and ER-risk (455 cases; 1554 controls)

Number of RoHs	OR	95% CI	P
6-18	1.00	Ref	
19-21	0.79	0.60-1.05	0.10
22-24	0.73	0.54-0.98	0.04
>25	0.93	0.69-1.25	0.64
Total Length (Mb)			
7.5-24.1	1.00	Ref	
24.2-28.6	0.67	0.49-0.89	0.007
28.7-33.3	0.68	0.51-0.91	0.01
>33.5	0.83	0.62-1.12	0.23

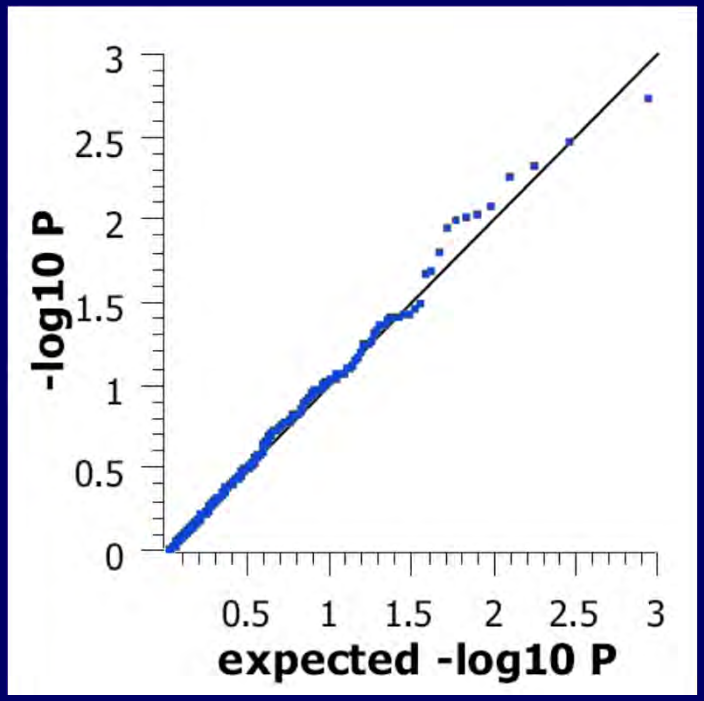
Locus-by-locus association analyses

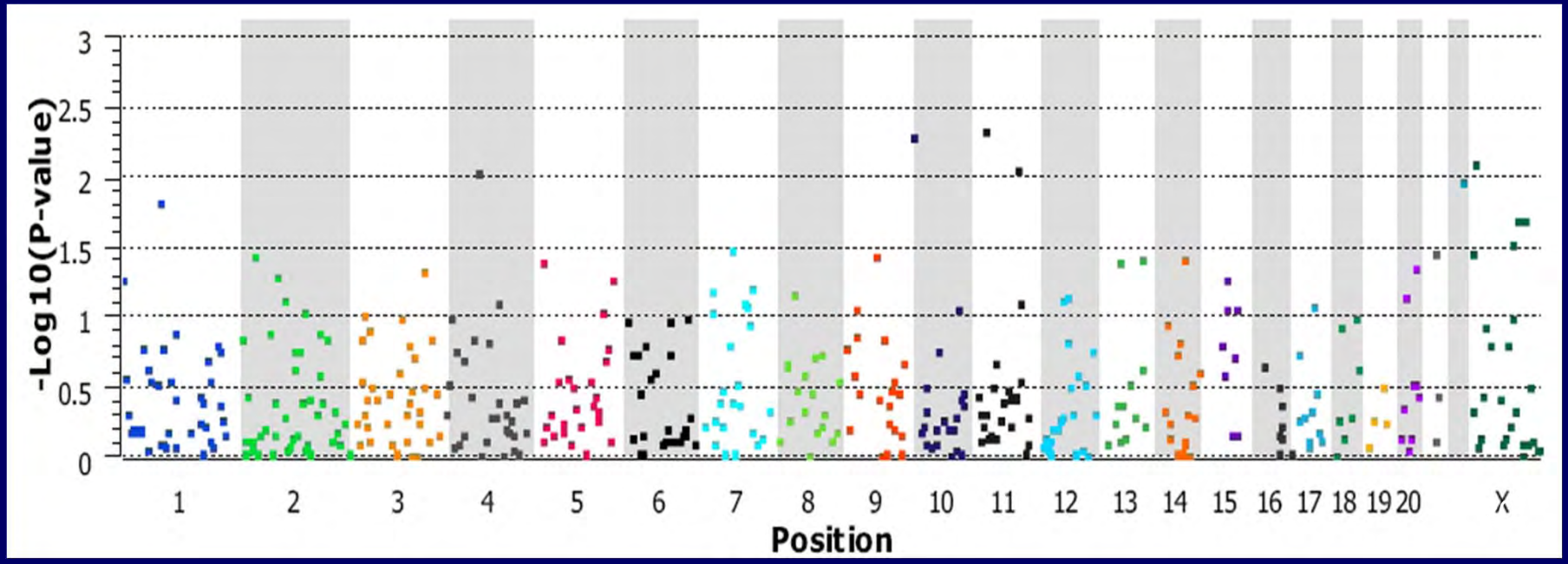
- Test association between breast cancer and
  - Binary RoH status at each “common” RoH
  - “Percent coverage” of total length of RoH
- Using logistic regression adjusted for
  - Age at diagnosis/interview
  - PCA-derived ancestry (5 PCs)
- Examine by ER status
  - 456 ER- and 975 ER+

Association Results

(1641 cases; 1554 controls)

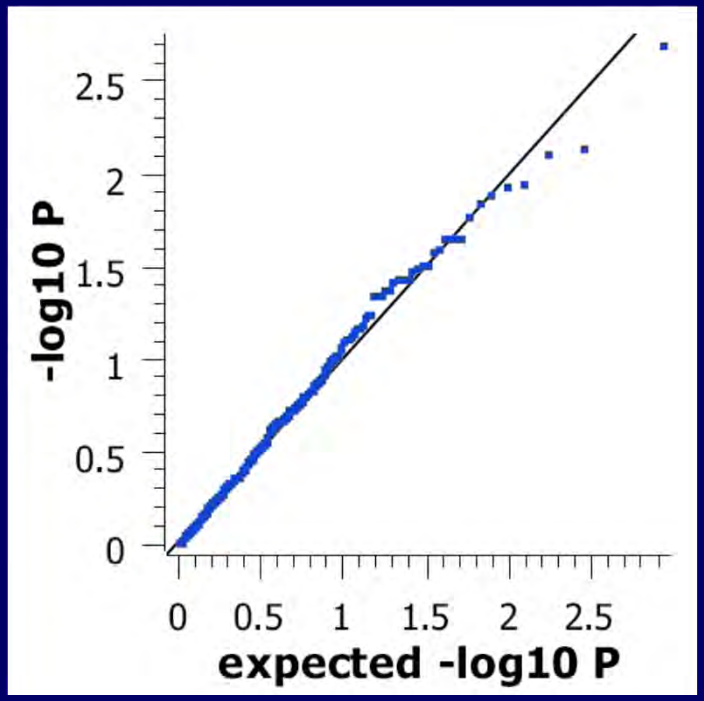
- P-values do not strongly deviate from their expected distribution under the null
- Strongest associations observed on chromosomes 11 (P=0.005) and 10 (P=0.005)

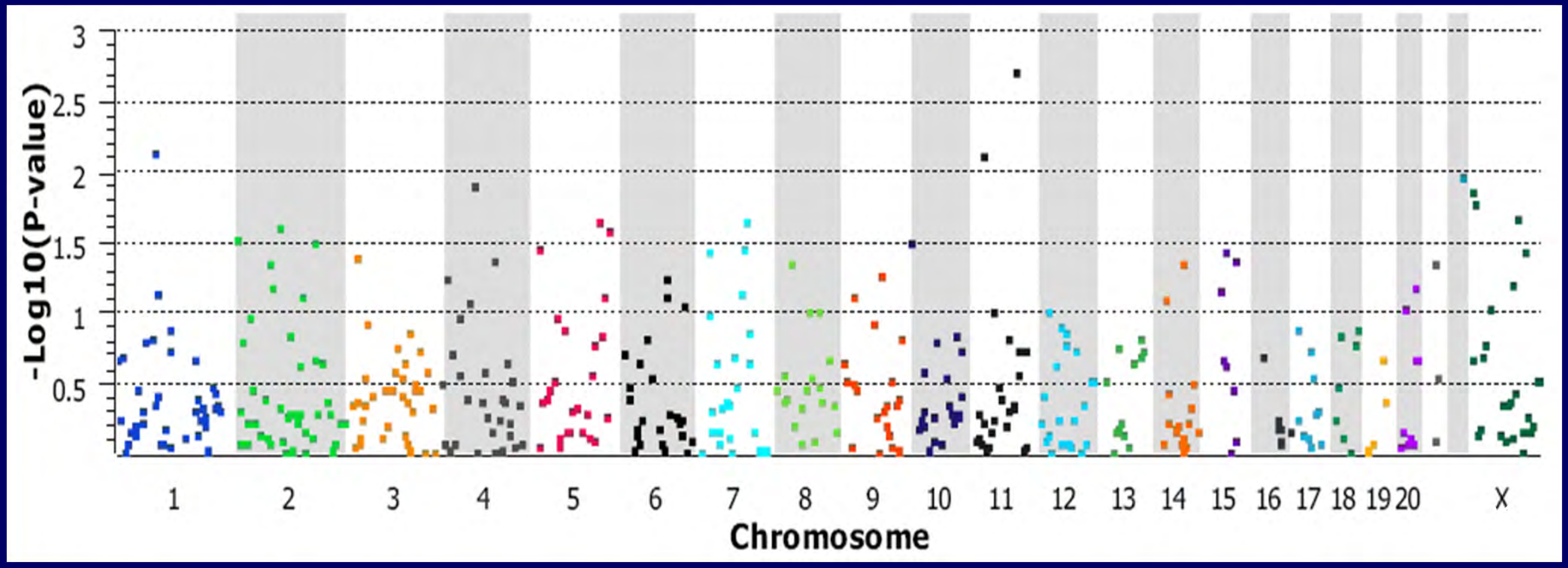




Association Results (942 ER+ cases)

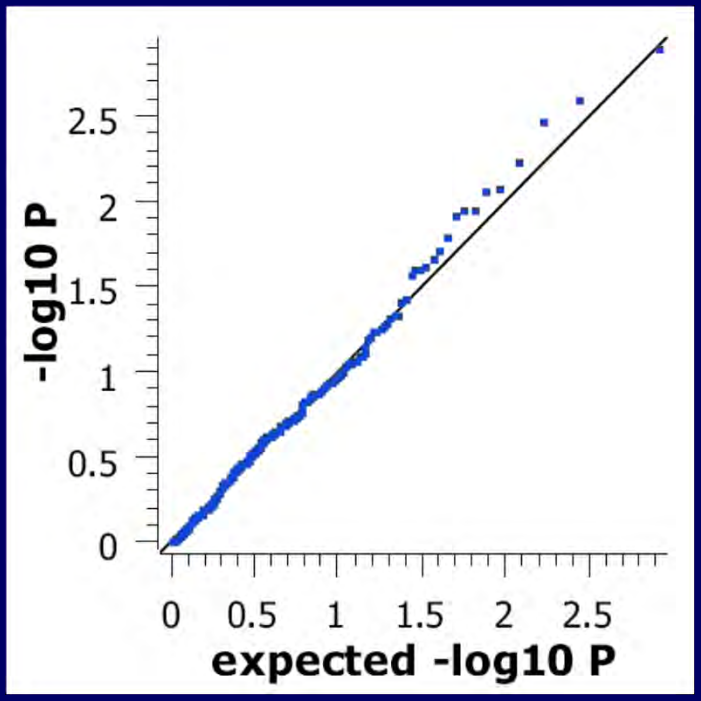
- P-values do not strongly deviate from their expected distribution under the null
- Strongest associations observed on chromosomes 11 (P=0.002) and 1 (P=0.007)

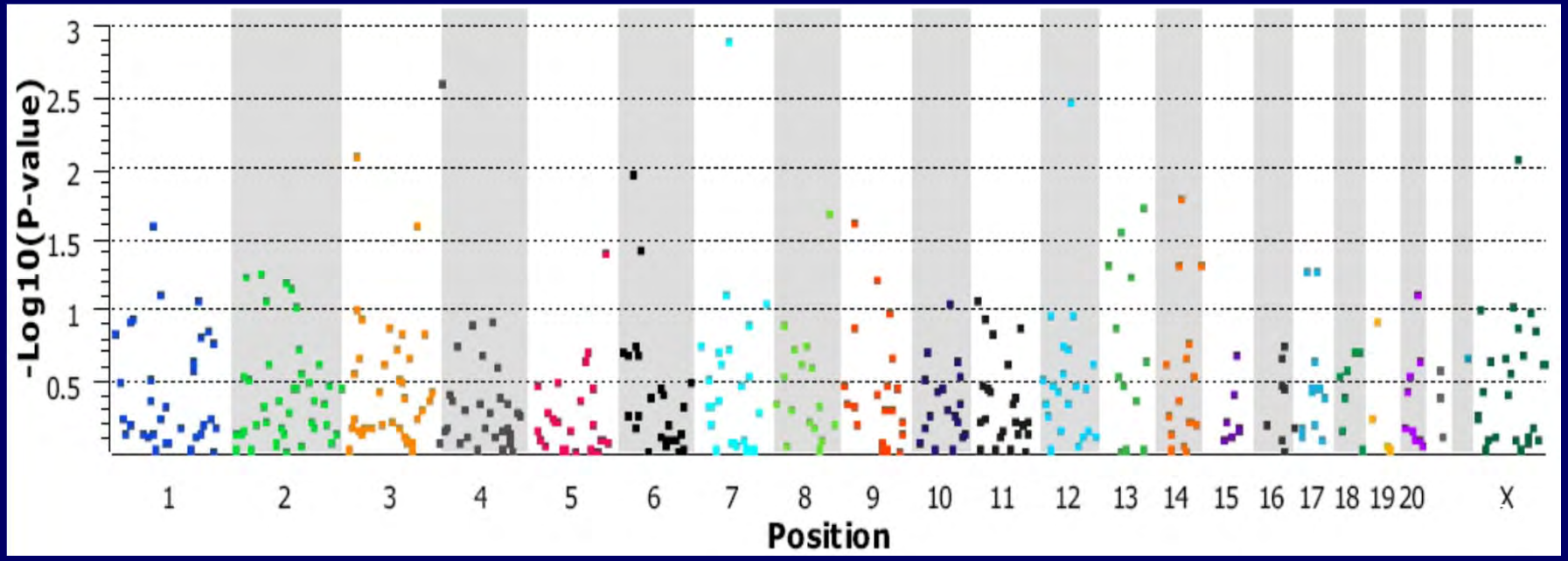




Association Results (455 ER- cases)

- P-values do not strongly deviate from their expected distribution under the null
- Strongest associations observed on chromosomes 7 (P=0.001) and 4 (P=0.003)





Association Results (PLINK method)

Number of RoHs	OR	95% CI	P
7-18	1.00	Ref	
19-21	1.04	0.86-1.26	0.68
22-24	0.95	0.78-1.16	0.65
>25	1.08	0.86-1.31	0.46
Total Length (Mb)			
8.8-27.9	1.00	Ref	
28.0-33.2	1.01	0.83-1.24	0.86
33.3-39.2	1.12	0.92-1.37	0.26
>39.3	1.12	0.92-1.37	0.28

Association Results (ER+)

Number of RoHs	OR	95% CI	P
7-18	1.00	Ref	
19-21	1.11	0.90-1.37	0.35
22-24	1.02	0.82-1.26	0.89
>25	1.15	0.93-1.42	0.21
Total Length (Mb)			
8.8-27.9	1.00	Ref	
28.0-33.2	1.07	0.86-1.33	0.54
33.3-39.2	1.17	0.94-1.46	0.17
>39.3	1.21	0.97-1.51	0.09

Association Results (ER-)

Number of RoHs	OR	95% CI	P
7-18	1.00	Ref	
19-21	0.92	0.71-1.76	0.49
22-24	0.83	0.64-1.08	0.16
>25	0.95	0.76-1.26	0.87
Total Length (Mb)			
8.8-27.9	1.00	Ref	
28.0-33.2	0.92	0.71-1.18	0.50
33.3-39.2	1.04	0.80-1.34	0.77
>39.3	0.99	0.76-1.29	0.94

Comparison with prior study of ER+ (top 6 signals)

Enciso-Mora	**Pierce**
10q21.2	11q21-11q22.1
5q15-5q21.2	11p14.3
6q22.31-6q22.33	1p31.1
3p22.2	4q13.3
3q21.2	22q13.1
1p31.3	Xp22.2

Conclusions

- Little evidence that overall RoH content or specific RoHs contribute to early-onset breast cancer risk
  - Suggestive association for total Mb of RoH and ER+
- The utility of RoH mapping for detection of cancer susceptibility loci appears to be somewhat limited
  - Underpowered in standard GWA studies?
  - Appropriate for specific populations?
- However, with available of large GWAS datasets associations for RoHs can be assessed for other cancer phenotypes